Teaching Machines to Extract Main Content for Machine Reading Comprehension

Zhaohui Li[1], **Yue Feng**[1], **Jun Xu**[2,3,*], **Jiafeng Guo**[1], **Yanyan Lan**[1], **Xueqi Cheng**[1]
[1]CAS Key Lab of Network Data Science and Technology
Institute of Computing Technology, Chinese Academy of Sciences
[2]Beijing Key Laboratory of Big Data Management and Analysis Methods
[3]School of Information, Renmin University of China
{lizhaohui,fengyue}@software.ict.ac.cn, junxu@ruc.edu.cn, {lanyanyan,guojiafeng,cxq}@ict.ac.cn

Abstract

Machine reading comprehension, whose goal is to find answers from the candidate passages for a given question, has attracted a lot of research efforts in recent years. One of the key challenge in machine reading comprehension is how to identify the *main content* from a large, redundant, and overlapping set of candidate sentences. In this paper we propose to tackle the challenge with Markov Decision Process in which the main content identification is formalized as sequential decision making and each action corresponds to selecting a sentence. Policy gradient is used to learn the model parameters. Experimental results based on MSMARCO showed that the proposed model, called MC-MDP, can select high quality main contents and significantly improved the performances of answer span prediction.

Introduction

Machine reading comprehension (MRC) aims to extract answers from a set of passages according to a question. Given a question, a large number of candidate passages could be involved. Therefore, how to identify answer-related content becomes a critical issue. Existing approaches address the issue through ranking the passages or merging the candidate answers extracted from the passages (Wang et al. 2017). However, the issue is still far from being fully addressed due to: 1) ranking of the passages is not effective for MRC because the goal is to identify the right answer from related content, rather than to find and browse the relevant passages; 2) the candidate passages could be similar in content or contain some wrong information, which makes the extracted answers overlapping or contain unrelated information.

Ideally, an MRC model shall follow the RC process of human beings. That is, human beings may first read the given question, and then go through the whole passages to identify the question-related contents, and finally conclude the answer. The attention of human beings should focus on the question-related sentences during the whole process.

In this paper, we proposed a model that could automatically identify the main content from the candidate passages through mimicking the human PC process, on the basis of the attention mechanism (Seo et al. 2016) in deep learning and the Markov Decision Process (MDP) (Puterman 2014). Specifically, given the question and the candidate passages, the attention mechanism is employed to learn the question-aware embeddings for the question and every passage sentences. After that, the MDP scans all of the sentences and selects sentences as the main content. The model, referred to as MC-MDP, is trained with end-to-end manner and policy gradient is used to update the parameters.

It is shown that MC-MDP can identify the sentences that contain the answers with high precision, and thus has the ability of boosting the performances of span prediction models. Experimental results based on MSMARCO also showed that the combinations of MC-MDP and span prediction models can improve the MRC performances in terms of F1 and Rouge-L.

Our Approach: MC-MDP

MC-MDP consists of an encoding step which employs attention mechanism to obtain question-aware representations for the question and sentences, and a MDP step for constructing the main content sequence.

Question-awared representation

Following the practices (Seo et al. 2016), two bi-direction attention models (two BiLSTMs) are respectively employed to learn the representations of the question (represented as $\mathbf{q}$), and the sentences (represented as $X = \{\mathbf{x}_1, \cdots, \mathbf{x}_M\}$) in the passages, where M is the total number of sentences in all of the passages related to the question.

Main content construction

The learned representations are used as the inputs to the following MDP model, whose goal is to construct an agent for selecting a set of sentences as the main content. The elements of the MDP is defined as:

States $\mathcal{S}$: State at step t is a triple $s_t = [\mathbf{Q}, \mathcal{Z}_t, X_t]$, where $\mathbf{Q}$ is the question, $\mathcal{Z}_t = \{\mathbf{x}_{(n)}\}_{n=1}^{t}$ is the sequence of t selected sentences, and X_t is the set of remaining sentences. At the beginning ($t = 0$), the state is initialized as $s_0 = [\mathbf{Q}, \emptyset, X]$, where X contains the top w sentences in all the candidate passages. Note the agent will scan all sentences one by one and the w-size window is used to control the number of actions the agent can choose from.

Actions $\mathcal{A}$: At each time step t, the $\mathcal{A}(s_t)$ is the set of actions the agent can choose, each corresponds to a sentence in X_t. That is, action $a_t \in \mathcal{A}(s_t)$ selects a sentence $\mathbf{x}_{m(a_t)} \in$

[*] Corresponding author: Jun Xu

Algorithm 1 MC-MDP Training

```
Input: Training set $D = \{(\mathbf{Q}^{(n)}, X^{(n)}, A^{(n)})\}_{n=1}^{N}$ and
    learning rate $\eta$
1: Initialize parameters $\mathbf{\Theta} \leftarrow$ random values in $[-1, 1]$
2: repeat
3:   for all $(\mathbf{Q}, X, A) \in D$ do
4:     Sample $(\mathbf{s_0}, \mathbf{a_0}, r_1, \cdots, \mathbf{s_{M-1}}, \mathbf{a_{M-1}}, r_M) \sim \mathbf{p}$
5:     for $t = 0$ to $M - 1$ do
6:       $G_t \leftarrow \sum_{k=0}^{M-1-t} r_{t+k+1}$
7:       $\mathbf{\Theta} \leftarrow \mathbf{\Theta} - \eta G_t \nabla_{\Theta} \log p(a_t|s_t; \Theta)$
8:     end for
9:   end for
10: until converge
```

X_t as the main content sequence, where $m(a_t)$ is the index of the sentence corresponding to a_t.

Transition T**:** The function $T : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$ is defined as two steps: 1) append sentence $\mathbf{x}_{m(a_t)}$ to $\mathcal{Z}_t$; 2) Set X_{t+1} as $\{\mathbf{x}_{m(a_t)+1}, \cdots, \mathbf{x}_{m(a_t)+w}\}$, where w is the window size.

Reward R**:** The reward function at t step is defined as the arithmetic mean of $F_1(t)$ and $\mathrm{Rouge-L}(t)$.

Policy $\mathbf{p}$**:** Each probability in the policy is a normalized function whose input is the bilinear product of the LSTM and the selected sentence:

$$p(a|s) = \frac{\exp\left\{\mathbf{x}_{m(a)}^{T} \mathbf{U}_p \, \mathrm{LSTM}(s)\right\}}{\sum_{a' \in \mathcal{A}(s)} \exp\left\{\mathbf{x}_{m(a')}^{T} \mathbf{U}_p \, \mathrm{LSTM}(s)\right\}}. \quad (1)$$

Learning with policy gradient

The model has parameters to learned and Algorithm 1 shows the procedure. At each iteration, for each training instances, an episode $E = (\mathbf{s_0}, \mathbf{a_0}, \mathbf{r_1}, \cdots, \mathbf{s_{M-1}}, \mathbf{a_{M-1}}, \mathbf{r_M})$ is sampled according to Equation (1). After that, the long-term return G_t, which is the discounted sum of rewards from position t, is calculated and gradient is then estimated to update parameters, as shown in (line 6 and 7 of Algorithm 1. Note that the sequential sentence selecting procedure in Algorithm 1 makes it possible to filter out the overlapping sentences as the chosen of the actions in one state depends on its preceding actions. Another merits of MC-MDP is that the question/sentence representations and the MDP policy can be trained jointly and enjoy the end-to-end training of the model. At the on-line time, the agent selects the sentences from the passages according to the learned representations and the policy function.

Experiment and Analysis

The proposed MC-MDP model was tested on the public available benchmark MSMARCO, which consists of 100K questions and 1M passages. TextBlob was used to conduct the preprocessing and each word was represented as a 300-dimensional vector by GloVe. Match-LSTM (Wang and Jiang 2016) and BiDAF, two popular span prediction models, were employed to extract the final answers. As for baseline, we choose the GP (Golden Passage), which chooses passage has the largest overlap with the question as the main content. We also tested the performances of Match-LSTM and BiDAF without any main content. The evaluation metrics include ROUGE-L and BLEU-1.

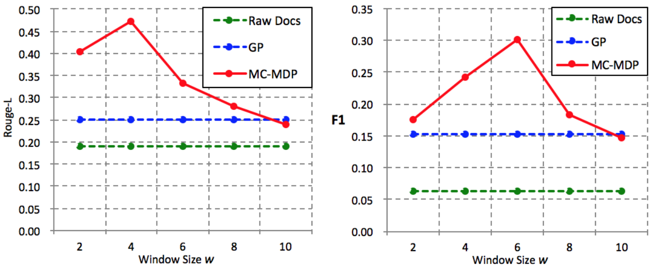

Figure 1: Scores of experiments with different methods

Table 1: Comparison of results in verification tests.

Method	ROUGE-L%	BLEU-4%
Match-LSTM	30.67	31.05
BiDAF	31.52	32.11
GP + Match-Lstm	37.33	40.72
GP + BiDAF	37.56	41.24
MC-MDP + Match-LSTM	**43.97**	**43.23**
MC-MDP + BiDAF	**44.12**	**43.45**

MC-MDP has some parameters. The dimensions of all hidden layers were set to 150 and the dropout rate between layers was set to 0.8. The results reported in Table 1 show that MC-MDP outperformed all of the baselines, indicating the effectiveness of MC-MDP in selecting the main content. We also tested the impact of window size w in MC-MDP, shown in Figure 1. We can see that MC-MDP achieved the best performances when w was set as 5.

Conclusion

In this paper we proposes to identify the main content of a question for machine reading comprehension. The model, called MC-MDP, first learn the question-aware representations for the sentences with the attention mechanism. It then selects the sentences as the main content with an MDP. Reinforcement learning algorithm was proposed to learn the model parameters in an end-to-end manner. Experiments show that MC-MDP can outperform the baselines through selecting the sentences containing the answers with high precision.

Acknowledgement

This work was funded by the National Natural Science Foundation of China (NSFC) under Grants No.61872338, 61773362, 61425016, 61472401 and 61722211, the Youth Innovation Promotion Association CAS under Grants No. 20144310, and 2016102, and the National Key R&D Program of China under Grants No. 2016QY02D0405.

References

Puterman, M. L. 2014. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Seo, M.; Kembhavi, A.; Farhadi, A.; and Hajishirzi, H. 2016. Bidirectional attention flow for machine comprehension. *arXiv preprint arXiv:1611.01603*.

Wang, S., and Jiang, J. 2016. Machine comprehension using match-lstm and answer pointer. *arXiv preprint arXiv:1608.07905*.

Wang, S.; Yu, M.; Guo, X.; Wang, Z.; Klinger, T.; Zhang, W.; Chang, S.; Tesauro, G.; Zhou, B.; and Jiang, J. 2017. Reinforced reader-ranker for open-domain question answering. *arXiv preprint arXiv:1709.00023*.